

**计算机网络**

**课程实验报告**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 实验名称 | HTTP 代理服务器的设计与实现 | | | | | |
| 姓名 | 石翔宇 | | 院系 | 计算学部 | | |
| 班级 | 1903103 | | 学号 | 1190200523 | | |
| 任课教师 | 刘亚维 | | 指导教师 | 刘亚维 | | |
| 实验地点 | 格物207 | | 实验时间 | 2021.10.31 | | |
| 实验课表现 | 出勤、表现得分(10) |  | 实验报告  得分(40) |  | 实验总分 |  |
| 操作结果得分(50) |  |
| 教师评语 | | | | | | |
|  | | | | | | |

****

|  |
| --- |
| 实验目的： |
| 熟悉并掌握 Socket 网络编程的过程与技术；深入理解 HTTP 协议，掌握 HTTP 代理服务器的基本工作原理；掌握 HTTP 代理服务器设计与编程实现的基本技能。 |
| 实验内容： |
| (1) 设计并实现一个基本 HTTP 代理服务器。要求在指定端口（例如8080）接收来自客户的 HTTP 请求并且根据其中的 URL 地址访问该地址所指向的 HTTP 服务器（原服务器），接收 HTTP 服务器的响应报文，并将响应报文转发给对应的客户进行浏览。  (2) 设计并实现一个支持 Cache 功能的 HTTP 代理服务器。要求能缓存原服务器响应的对象，并能够通过修改请求报文（添加 if-modified-since头行），向原服务器确认缓存对象是否是最新版本。（选作内容，加分项目，可以当堂完成或课下完成）  (3) 扩展 HTTP 代理服务器，支持如下功能：（选作内容，加分项目，可以当堂完成或课下完成）  a) 网站过滤：允许/不允许访问某些网站；  b) 用户过滤：支持/不支持某些用户访问外部网站；  c) 网站引导：将用户对某个网站的访问引导至一个模拟网站（钓鱼）。 |
| 实验过程： |
| **HTTP代理服务器的实现流程：**   1. 初始化服务器端Socket并开启监听； 2. 等待客户端链接，开启新线程并创建客户端的Socket； 3. 用户过滤，若用户IP在黑名单内，则禁止访问； 4. 解析客户的HTTP请求报文获得目的Host的URL，IP地址； 5. 网站过滤，若URL在过滤列表中，则返回错误； 6. 网站引导，若URL在引导列表中，将URL和IP地址换成引导目的地址； 7. 连接目的Host获得Host端的Socket； 8. 在缓存中查找目的Host：    1. 若没有找到，则直接将客户的HTTP请求报文发送给目的Host，将目的Host返回的报文发送给客户。    2. 否则，在客户的HTTP请求报文中加入if-modified-since头行，发送给目的Host，若返回304则将缓存中的内容发送给客户；若返回200则将返回的报文发送给客户。 9. 按照目的Host返回的报文更新缓存。   **流程图：**    **关键功能实现：**  解析客户的HTTP请求报文：  首先从第一行中找到HTTP请求类型，目前只支持GET和POST，同时提取请求的URL。再利用strtok\_s函数按照\r\n将报文中的不同行分开，分别浏览每一行，找到有Host的行，将Host提取出来。    **用户过滤：**  disabledUser数组中存放着被禁止的用户IP。在获得用户的socket后可以得知用户的IP地址，将其与数组中的IP进行匹配，若匹配成功则返回USER\_BLOCKED表示该用户已被禁止访问。    **网站过滤：**  disabledHost数组中存放着被禁止的网站Host。在解析客户的HTTP请求报文后，可以获得该用户想要访问的网站Host，将其与数组中的Host进行匹配，若匹配成功则返回HOST\_BLOCKED表示该网站已被禁止访问。    **网站引导：**  induceSites数组中存放着被引导的网站Host，targetSites数组中存放着引导目标网站Host，两者一一对应。在解析客户的HTTP请求报文后，可以获得该用户想要访问的网站Host，首先通过网站过滤功能，若不是被过滤的网站，则进入网站引导。将其与induceSites数组中的内容进行匹配，若匹配成功，则将其更改为induceSites数组中对应的Host。    **缓存：**  我们首先定义关于网页缓存的类，内容包括HTTP头，网页缓存，最后更新时间，网页缓存长度。    缓存实现的第一步需要在内存中查找缓存。若当前请求的HTTP头中的方法、URL与Host都与某缓存相匹配的话，则缓存命中，直接返回该缓存。    接下来需要向服务器发送带If-Modified-Since的报文。我们实现的方法是在用户发来的请求报文中查找Host字段，在该字段前面插入If-Modified-Since字段。    若服务器返回304则说明网站内容未更新，直接将缓存中内容返回给用户；否则将服务器返回的报文返回给用户。  最后要做的就是更新缓存内容。首先将网站最后更新时间解析出来，再将最后更新时间和内容更新给缓存。若缓存中本来就没有该网站的内容，则还需要新建缓存。 |
| 实验结果： |
| **基础功能：**  首先是代理服务器最基础的功能实现。我们访问<http://cs.hit.edu.cn/>结果如下：    代理服务器端显示日志如下：    **用户过滤：**  我们将本地IP地址127.0.0.1限制访问，结果如下：    可以看到当前用户已经无法访问任何网站。代理服务器端显示日志如下：    **网站过滤：**  我们将网站[jwc.hit.edu.cn](http://jwc.hit.edu.cn/)限制访问，结果如下：    可以看到已经无法访问jwc.hit.edu.cn。代理服务器端显示日志如下：    从日志可以看到发生了错误，编号为5，这正是网站被过滤时提示的错误编号。    **网站引导：**  我们将对网站today.hit.edu.cn的所有访问都转到jwts.hit.edu.cn中，结果如下：    可以看到地址栏中写的是today.hit.edu.cn但显示的却是jwts.hit.edu.cn的内容。代理服务器端显示日志如下：    **缓存：**  我们将两次访问cs.hit.edu.cn来测试缓存功能，代理服务器端显示日志如下：  这是第一次访问cs.hit.edu.cn中某个jpg图片资源时的日志：    这是第二次访问的日志：    可以看到缓存的长度增加了（还有别的资源也被缓存下来），并且在第二次访问时代理服务器端给出了日志Using cache，这表明代理服务器判断缓存无更新，将缓存里的内容返回给了用户。 |
| 问题讨论： |
| 在实现网站诱导时，发现不更改HTTP报文中的Host和URL字段也能返回目标诱导网站的页面，但只能是主页面，其下的URL是无法正常访问的。 |
| 心得体会： |
| 通过此次实验，我更加深刻地理解HTTP协议的原理以及工作方式，掌握了Socket网络编程技术，了解了HTTP代理服务器的基本原理，实现了一个小的HTTP代理服务器，激发了我对计算机网络的兴趣。 |